Problem statement

We all want to do accurate Bayesian inference quickly:

- In terms of compute (wall time, model evaluations, parallelism)
- In terms of analyst effort (tuning, algorithmic complexity)

Markov Chain Monte Carlo (MCMC) can be straightforward and accurate but slow.

Black Box Variational Inference (BBVI) can be faster alternative to MCMC. **But...**

- BBVI is cast as an optimization problem with an intractable objective $\Rightarrow$
- Most BBVI methods use **stochastic gradient** (SG) optimization $\Rightarrow$
  - SG algorithms can be hard to tune
  - Assessing convergence and stochastic error can be difficult
  - SG optimization can perform worse than second-order methods on tractable objectives
- Many BBVI methods employ a **mean-field (MF) approximation** $\Rightarrow$
  - Posterior variances are poorly estimated

Our proposal: replace the intractable BBVI objective with a fixed approximation.

- Better optimization methods can be used (e.g. true second-order methods)
- Convergence and approximation error can be assessed directly
- Can correct posterior covariances with linear response covariances
- This technique is well-studied (but there's still work to do in the context of BBVI)

$\Rightarrow$ **Simpler, faster, and better BBVI posterior approximations ... in some cases.**

Outline

- BBVI Background and our proposal
  - Automatic differentiation variational inference (ADVI) (a BBVI method)
  - Our approximation: "Deterministic ADVI" (DADVI)
  - Linear response (LR) covariances
  - Estimating approximation error
- Experimental results: DADVI vs ADVI
  - DADVI converges faster than ADVI, and requires no tuning
  - DADVI's posterior mean estimates' accuracy are comparable to ADVI
  - DADVI+LR provides more accurate posterior variance estimates than ADVI
  - DADVI provides accurate estimates of its own approximation error
  - ADVI often results in better objective function values (eventually)
- Why don't we do DADVI all the time?
  - DADVI fails for expressive BBVI approximations (e.g. full-rank ADVI)
  - Pessimistic dimension dependence results from optimization theory
  - ...which may not apply in certain BBVI settings.